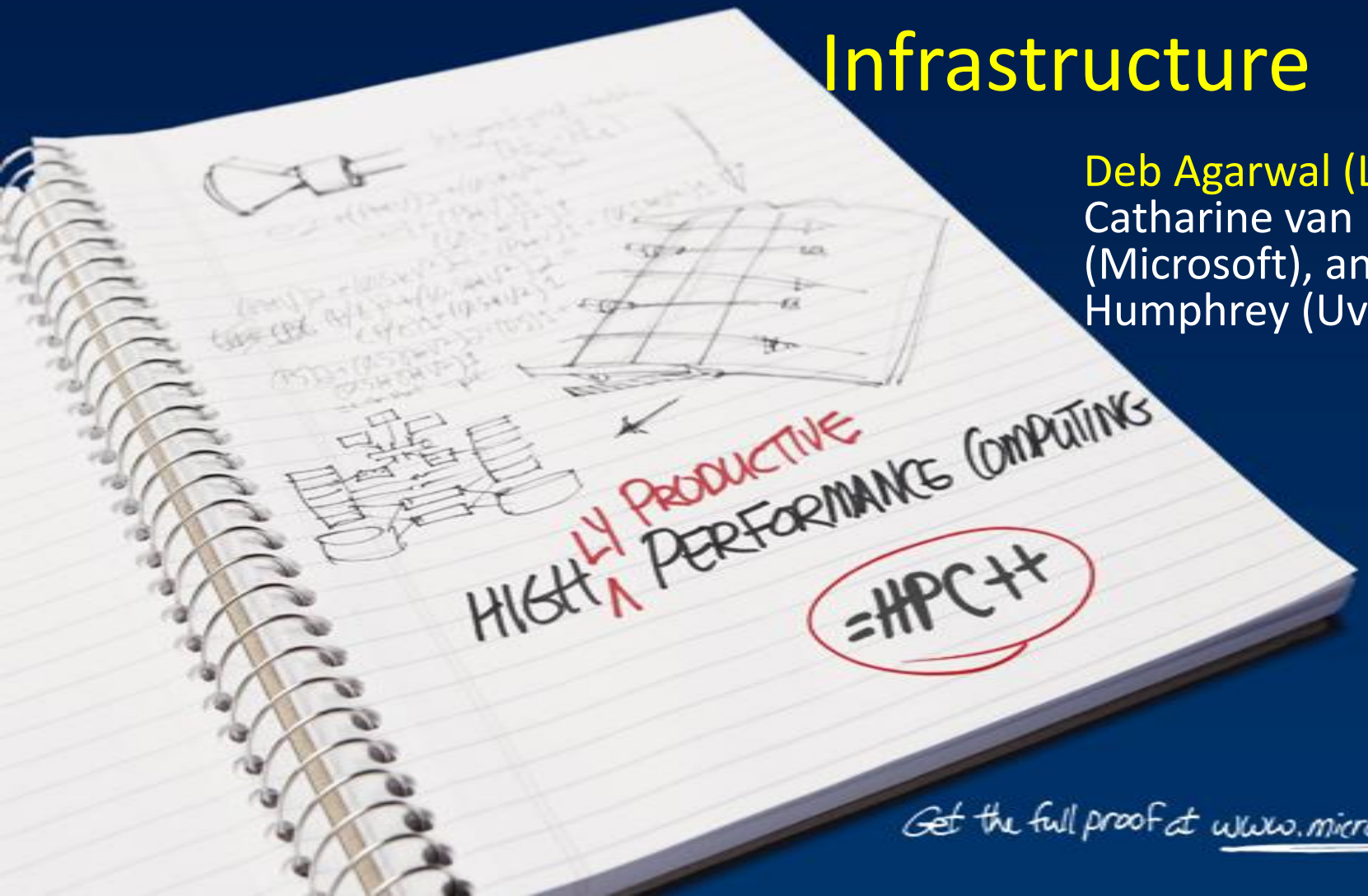
Windows HPC Server 2008

Microsoft

Berkeley Water Center Eco-Science Infrastructure

Deb Agarwal (LBL), Catharine van Ingen (Microsoft), and Marty Humphrey (Uva)

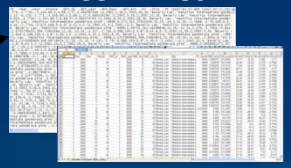


Get the full proof at www.microsoft.com/hpc

Ecological Data

- Data comes from many sources
  - Multiple entities fly satellites and operate fixed sensors – imagery from NASA, precipitation from NOAA, discharge from USGS, water quality from EPA
  - Scientists take data in field campaigns, sensor or other instrument deployments, photographs
  - Collecting site properties such as channel cross-section or leaf area index done by individuals as well as agencies
- Data repositories not universal
  - passed around in e-mail
  - edited at will
  - silently corrected in repositories when "everybody knows"
- Data comes from unexpected places
  - Published literature supplies various numbers and graphs giving context, sanity checks on the results
- Analysis routines usually custom

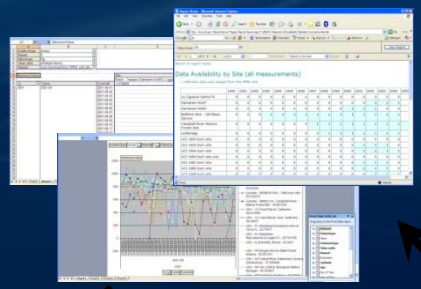
Agency web sites, streaming sensor data, or other source

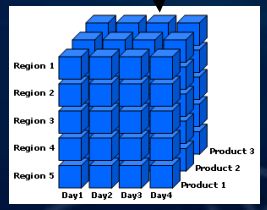


CSV Files

BWC SQL Server Database

Reports, Excel Pivot Table, MatLab, ArcGIS





BWC Data Cube

Example Dataset - FLUXNET

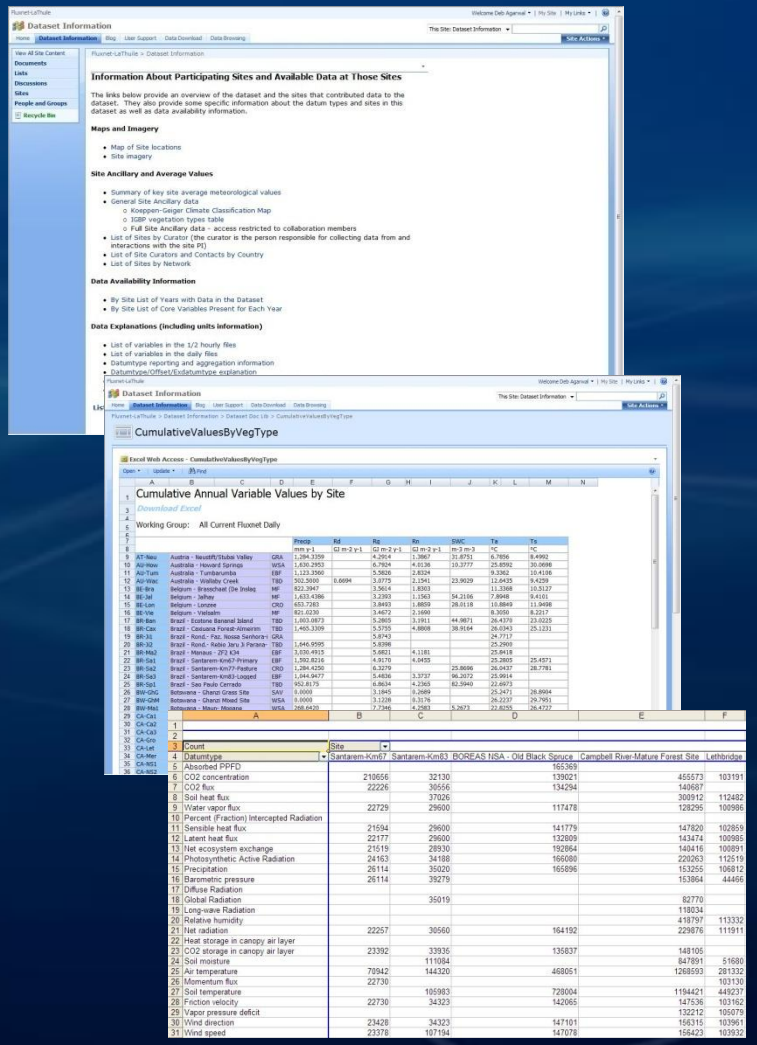
Windows HPC Server 2008

- Global carbon flux and micrometeorological measurements
  - 960 site years of data
    - 40 variables and 60 quality tags
    - 1/2 hourly measurements
  - 253 sites
  - ~120 site investigators (collecting data)
  - 72 synthesis paper proposals involving ~110 collaborators
- Goal: Understand the role of photosynthesis in carbon cycle

Developing Common Infrastructure

Windows HPC Server 2008

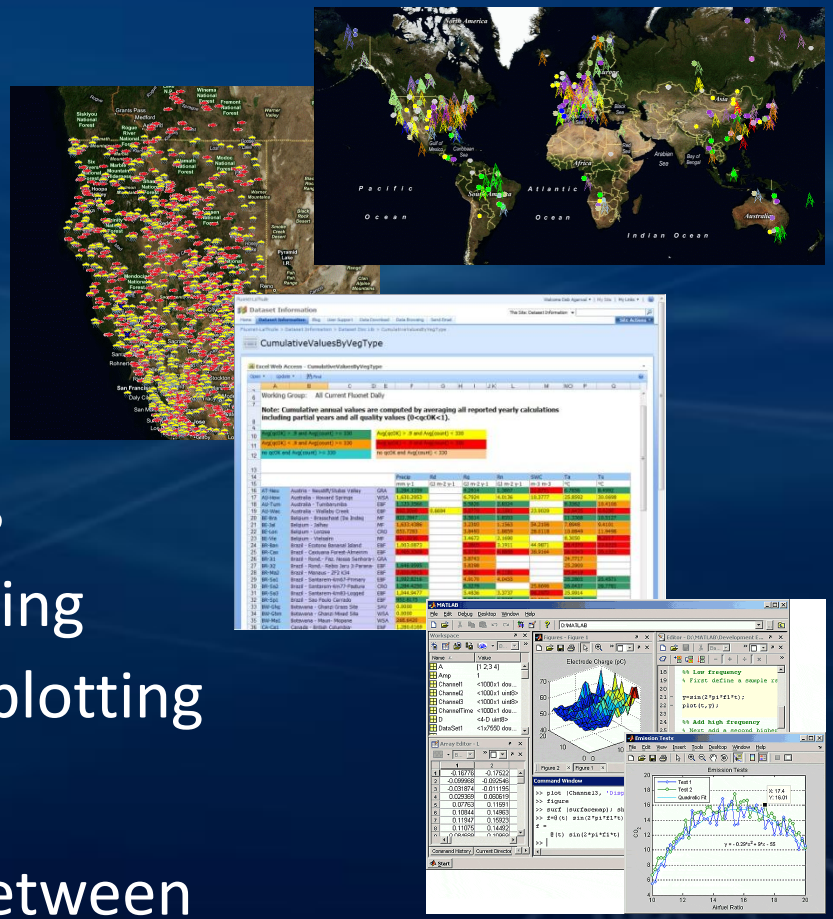
- Data ingested into a database
- Database schema organized by dataset, site, time, and variable
- On-line analytical processing (OLAP) data cubes provide browsing access
- Reports provide web pages describing the data
- Ancillary data gathered and curated
- Collaboration tools facilitate communication between data providers and data users as well as between data users



Interaction and Visualization

Windows HPC Server 2008

- Mashup maps
  - locate sensors
- Data summary reports
  - Data browsing
  - Data selection
  - Cross-site information
- OLAP data cube interfaces
  - Matlab – analysis and plotting
  - Excel – data browsing and plotting
- Collaboration capabilities
  - Improve communication between data producers and data consumers



Leveraging a Common Approach and Technology Base

Windows HPC Server 2008

Established Projects

- Russian River
  - Digital watershed built from existing data to look at changes over time affecting salmon
- Carbon Flux Measurement Network
  - Regional (AmeriFlux) and global (FLUXNET) datasets compiled from local carbon flux measurements
- Pajaro River
  - Digital watershed built from existing data to look at natural channel flood control alternatives (listed in top 10 “endangered river” in 2006)

Projects Getting Underway Now

- National Marine Fisheries Service
  - Digital watershed to support fish recovery planning for northern California
- California watersheds server
- North American Carbon Program – Carbon flux modeling

Acknowledgements

Berkeley Water Center, University of California, Berkeley, Lawrence Berkeley Laboratory
- Jim Hunt (UCB)
- Deb Agarwal (LBL)
- Robin Weber (UCB)
- Monte Good (LBL)
- Rebecca Leonardson (UCB student)
- Nick Butler (UCB student)
- Matt Rodriguez (SDSC Student)
- Carolyn Remick (UCB)
- Susan Hubbard (LBL)

University of Virginia
- Marty Humphrey
- Norm Beekwilder

Microsoft
- Catharine van Ingen
- Bora Beran
- Jayant Gupchup (student)
- Savas Parastidis
- Andy Sterland
- Nolan Li (student)
- Tony Hey
- Dan Fay
- Jing De Jong-Chen
- Stuart Ozer
- SQL product team
- Jim Gray

Ameriflux Collaboration
- Beverly Law (UOregon)
- Youngryel Ryu (UCB postdoc)
- Tara Stiefl (UO student)
- Chris Williamson (UO student)
- Gretchen Miller (UCB student)
- Tom Boden (ORNL)

North American Carbon Program
- Peter Thornton (ORNL)
- Kevin Schaefer (NSIDC)
- Daniel Ricciuto (ORNL)

FLUXNET Collaboration
- Dennis Baldocchi (UCB)
- Rodrigo Vargas (UCB postdoc)
- Dario Papale (UTuscia)
- Markus Reichstein (Max Plank)
- Bob Cook (ORNL)
- Dorothea Frank (Max Plank)

National Marine Fisheries Service
- Charlotte Ambrose (NOAA)
- Pat Stiefer (SEC)

Sonoma County Water Agency
- Chris Delaney

The saga continues at: http://bwc.berkeley.edu and http://www.fluxdata.org

BERKELEY LAB

UNIVERSITY of VIRGINIA
ENGINEERING

Microsoft
Research

iLEAPS

Fluxnet Canada

Africa

CARBO
EUROPE

AsiaFlux

NSF

LBA
The Large Scale
Biosphere-Atmosphere
Experiment
in Amazônia

AmeriFlux

OzFlux

ChinaFLUX

KoFlux

NECC

Terrestrial Carbon Observation System
OS
SiBERIA

U.S. DEPARTMENT OF
ENERGY
Office of
Science